Early screening of autism spectrum disorder using cry features

Aida Khozaei[1], Hadi Moradi[1, 2], Reshad Hosseini[1], Hamidreza Pouretemad[3], Bahareh Eskandari[3]

[1] School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran.

[2]Adjunct Research Professor, Intelligent Systems Research Institute, SKKU, Suwon, South Korea

[3] Department of Psychology, Shahid Beheshti University, Tehran, Iran

Abstract

Due to the importance of automatic and early autism screening, in this paper, a cry-based screening approach for children with Autism Spectrum Disorder (ASD) is introduced. During the study, we realized that the ASD specific features are not necessarily observable among all ASD subjects and among all instances of each subject. Therefore, we proposed a new classification approach to be able to find such features and their correspondign instances. We tested the proposed approach and found two features that can be used to distinguish groups of children with ASD from Typically Developed (TD) children. In other words, these features are present in subsets of children with ASD not all of them. The approach has been tested on a dataset including 14 male and 7 female children with ASD and 14 male and 7 female TD children, between 18 to 53 months of age. The sensitivity, specificity, and precision of the proposed approach for male subjects were 85.71%, 100%, and 92.85%, respectively. These measures were 71.42%, 100%, and 85.71% for feamale subjects, respectively.

Keywords Screening, Autism spectrum disorder, Crying, Clustering, Classification, Feature Selection

Introduction

Children with Autism Spectrum Disorder (ASD) are defined by their abnormal or impaired development in social interaction and communication, as well as restricted and repetitive behaviors, interests, or activities [1]. The rapid growth of ASD in the past 20 years, has ignited many researches toward diagnosis and rehabilitation of ASD [2-5]. In the field of diagnosis, there are several well-established manual methods to diagnose children over 18 months [6]. However, due to lack of knowledge about ASD and lack of widespread expertness for screening and diagnosis, in practice, the average age of diagnosis is about 3 years [7]. Consequently, two main issues in ASD diagnosis are : 1) Early diagnosis that is crucial for effective therapy [7, 8] and 2) Automatic diagnosis/screening to reduce the cost and allow widespread screening by eliminating the need for experts.

Fortunately, there are works showing that the age of screening can even become lower than 18 months. For example, Thabtah and Peebles [9] reviewed several questionnaire-based approaches which may be able to screen ASDs above 6 months of age. However, these need trained practitioners to conduct the screening.

On the other hand, it is shown that fMRI [10] or EEG [11] can give discriminative features helping to diagnose ASD at earlier ages. Blood-based markers [12, 13] and prenatal immune markers [14] were also proposed to diagnose ASD that can be used right after birth. Although these approaches suggest new directions toward early ASD diagnosis, these are costly, need expertness, and are still at early stages of research and need further approval. Furthermore, approaches which involves methods such as fMRI or EEG, are hard to be used on children, especially children with autism.

There are studies which tried to propose non-invasive and cheaper methods for screening autism. For instance, Brisson et al. [15] observed voice features differences between children with ASD and Typically Developed (TD) children. Several studies, like [16], used speech-related features which are usable for screening of children older than 2. To reach the goal of early ASD screening, vocalizations of infants under 2 years of age were investigated in the literature [17-19]. Santos et al. [18] used vocalizations such as babbling to identify ASD children at the age of 18 months. Although Santos et al. [18] reported around 97% accuracy, it was obtained using k-fold cross-validation without considering subject-wise hold out to have unseen subjects in the test fold. The numbers of children with ASD and TD children in the study were 23 and 20 respectively. Oller et al. [17] included age in their classification approach and they excluded crying from vocalizations. They applied the method on 106 TD children and 77 children with ASD between 16 to 48 months of age and reached 86% accuracy. Pokorny et al. [19] extracted eGeMAPS parameter set [20], which includes 88 acoustic parameters, on 10 months old children (10 TD children and 10 children with ASD). This set consists of statistics calculated for 25 frequency-related, energy-related, and spectral low-level descriptors. They reached 75% accuracy.

Among the vocalizations, Esposito, Hiroi, and Scattoni [21] showed that cry is a promising biomarker for the identification of ASD children. Sheinkopf et al. [22] and Orlandi et al. [23] have shown that there are differences in the cry of children with ASD compared with TD chidlren. To the best of our knowledge, there is only one work that proposed a method for identifying ASD children using only cry [24]. They used a dataset of 5 children with ASD and 4 TD children older than two years of age. They extracted 187 sound features from which 55 features were selected, using forward selection. They reported 96.17% accuracy using k-fold cross validation without

considering subject-wise hold out which is a shortcoming of this work. In other words, it has been overfitted to the available data and may fail to correctly classify the new samples.

In all the above studies, it was assumed that the specific sound features, distinguishing children with ASD from TD children, are common among all the ASD cases. However, this may not be true for all the features. For instance, tip toe walking, which is one of the repetitive behaviors of children with ASD, appears in approximately 25% of these children [25]. In this paper, we propose a new cry-based identification approach for screening children with ASD. Our identification approach makes use of the assumption that all discriminative characteristics of autism do not appear in all ASD children. This assumption is in contrast with the assumption in the ordinary instance-based machine learning methods, which assumes that all instances of a class include all discriminative features needed for classification. In our proposed method, first discriminable instances of the cry, which exists in subsets of children with ASD, are found. Then it uses these instances to extract and select features to distinguish between these ASD instances from TD instances. It should be mentioned that the extracted features are age invariant and have been used to screen children with ASD between 18 to 53 months of age. This is in contrast to other approaches that either used a dataset of children with a specific age [18, 19] or used age information for classification [17]. The proposed approach has been implemented and tested on 62 subjects. The results show the effectiveness of the approach considering accuracy, sensitivity, and specificity of screening.

Method

In this section, we first introduce the subjects of this study details in the Subjects Subsection. Then in the Data collection and preprocessing Subsection, the protocol of data collection and

preprocessing is explained. In the Proposed subset instance classifier Subsection, our proposed classifier is introduced. The last Subsection is on the features extracted from the preprocessed data.

Subjects

There were 62 subjects, with at least one cry instance, i.e. a selected segment of the recorded cry voices, including 31 ASD and 31 TD. All subjects used in the training set were male, while in the test set 28 male and 14 female subjects between 18 and 53 months were included. Each set had equal numbers of TD and ASD classes. To eliminate the gender effects on the feature extraction and model training, these two set of subjects were evaluated separately. However, due to the lower number of female subjects with ASD, only a model was trained for screening male subjects. Nonetheless, the model was tested on both male and female subjects.

The inclusion criteria of the ASD subjects were those who had been just diagnosed as autistic, based on DSM V, with no other neurodevelopmental disorders such as hyperactivity. Furthermore, they had not taken any therapies or medication, or at most, they were in the first month of their therapy. There were two female subjects in the test set who had been diagnosed for more than a year. The diagnosis procedure included interviews with parents and clinical assessment of the child. At the first, parents completed Gilliam Autism Rating Scale-Second Edition (GARS-2) [26], then children were evaluated and observed by two PhD degree child clinical psychologist. In addition, diagnosis of ASD established by at least a child psychiatrist, separately, in a different setting. For full descriptions of these procedures, see our previous work [27]. TD children were selected from those in an age range similar to the ASD ones from volunteer families at homes and health centers. They had no official diagnosis of neurological or psychological disorder. All children were included in this study having their parents' consent. The children with ASD were

older than 20 months with the age average, standard deviation and, range of 35.6, 8.8, and 33 months, respectively. The TD children were younger than 51 months with the age average, standard deviation, and range of about 30.8, 10.3, and 33 months, respectively. It should be mentioned that the diagnosis of the children under 3 years of age was mainly based on experts' evaluation, not GARS score.

Tables 1 and 2 show the details of male and female subjects, respectively. In each table, the first column includes subject IDs. The second column contains ages of each subject in months. The number of instances of each subject and the total duration of all its instances in seconds is shown in columns 3 and 4, respectively. Since there were two device types used for recording, i.e. a high-quality recorder and typical cell phones, device type column shows which one was used for any subject. The next column includes GARS-2 scores of children with ASD. In 4 cases there were no GRAS grades available at the time of diagnosis. The column labeled as "Place" shows the location of recording which can be homes (H), health centers (C4, C5, and C6), and autism centers (C1, C2, and C3). There were a total number of 359 samples for all children. 53.44 % of the samples were from ASD children and 46.56 % were from TD children.

Two groups of 10 TD and 10 ASD children were selected for training the classifiers such that two groups were as balanced as possible with respect to the age and recording device. Thus, each child in the TD group had a corresponding child in the ASD group with about the same age. Furthermore, the number of subjects with high-quality recordings was 6 ASD and 4 TD. As a result of this balancing on our data, we obtained training subjects with the age between 20 and 51 months. The average ages in the training set were 32.7 and 35.2 months for ASD and TD subjects, respectively. The standard deviations are 9 and 9.9 months with the range of 25 and 30 months for ASD and TD subjects, respectively.

Table 1.The male subjects of the study.

		ID	Age (Month)	# of instances	Total Duration(S)	Device	GARS Grade	Place	Reason of cry
Children with autism	Train	ASD1	20	9	7.8	CP	104	C1	Annoyed/Uncomfortable
		ASD2	24	3	1.5	HQR	83	C2	Unwilling
		ASD3	26	5	2.1	HQR	120	C1	Annoyed/Uncomfortable
		ASD4	28	13	9.1	HQR	121	C2	Annoyed/Uncomfortable
		ASD5	29	14	26	HQR	89	C2	Unwilling/Complaining
		ASD6	31	4	2.4	HQR	87	C2	Unwilling/Complaining
		ASD7	36	11	11	HQR	87	C2	Unwilling/Complaining
		ASD8	43	2	0.7	CP	*	C2	Unwilling
		ASD9	45	3	2.6	CP	72	C2	Complaining
		ASD10	45	4	3.4	CP	*	H	Sleepy
	Test	ASD11	28	12	7.2	HQR	102	C2	Unwilling/ Uncomfortable
		ASD12	30	18	17.1	HQR	*	C3	Mother Separation
		ASD13	30	3	2.9	CP	*	H	Unwilling/Sleepy
		ASD14	31	5	2.3	HQR	73	C2/ H	Mother Separation/Hungriness
		ASD15	33	3	2.5	HQR	91	C2	Unwilling
		ASD16	33	2	2.5	HQR	104	C1	Annoyed/Uncomfortable
		ASD17	34	1	0.6	HQR	91	C2	Unwilling/Complaining
		ASD18	35	2	1.7	HQR	81	C1	Annoyed/Uncomfortable
		ASD19	37	1	0.6	HQR	94	C2	Unwilling/Complaining
		ASD20	40	19	14	HQR	91	C1	Annoyed
		ASD21	45	1	0.3	HQR	81	C2	Unwilling/Complaining
		ASD22	48	2	1.6	HQR	100	C2	Annoyed/Complaining
		ASD23	52	6	3.1	HQR	113	C2	Unwilling/Complaining
		ASD24	53	7	5.2	HQR	78	C1	Annoyed/Uncomfortable
Typically developed children	Train	TD1	21	11	14	HQR	NA	H	Complaining
		TD2	24	12	12	HQR	NA	C4	Scared/Unwilling
		TD3	26	2	2.3	HQR	NA	C5	Unwilling
		TD4	28	6	13	CP	NA	C5	Scared/Unwilling
		TD5	36	3	2.6	CP	NA	H	Unwilling/Complaining
		TD6	38	3	1.5	HQR	NA	C6	Complaining
		TD7	41	3	2.4	HQR	NA	H	Unwilling
		TD8	43	3	2.2	CP	NA	H	Sleepy
		TD9	44	2	1.2	CP	NA	H	Complaining
		TD10	51	2	1.7	CP	NA	H	Complaining
	Test	TD11	18	4	2	HQR	NA	C4	Scared
		TD12	18	7	5.1	HQR	NA	C4	Scared/Unwilling
		TD13	19	7	4.2	HQR	NA	C5	Unwilling
		TD14	20	9	8	HQR	NA	C5	Unwilling/Complaining
		TD15	21	4	1.2	HQR	NA	H	Complaining
		TD16	24	3	2.7	HQR	NA	C5	Scared /Unwilling
		TD17	24	2	1.5	HQR	NA	C5	Scared/Unwilling
		TD18	24	6	5.1	HQR	NA	C4	Unwilling/Complaining
		TD19	24	4	2.4	HQR	NA	C5	Unwilling/Complaining
		TD20	24	5	4.2	HQR	NA	C5	Unwilling/Complaining
		TD21	29	11	10	HQR	NA	H	Unwilling/Complaining
		TD22	30	4	2	HQR	NA	C5	Scared/Unwilling
		TD23	30	4	2	CP	NA	H	Unwilling
		TD24	43	12	11	HQR	NA	H	Complaining

There are two main sets of male subjects; train and test sets.

HQR High Quality Recorder, **CP** Cell Phone, **C1-C3** autism Centers, **C4-C6** health Centers, **H** Home, **NA** Not Applicable

*GARS grades at the time of diagnosis or starting the intervention was not available

Table 2.The female subjects of the study.

		ID	Age (Month)	# of instances	Total Duration(S)	Device	GARS Grade	Place	Reason of cry
ASD	Test	ASD25	25	12	14	HQR	85	C2	Unwilling/Complaining
		ASD26	26	5	2	CP	102	C1	Scared
		ASD27	31	3	1.7	HQR	94	C2	Unwilling/Complaining
		ASD28	32	2	1.3	HQR	100	C2	Unwilling/Complaining
		ASD29	41	8	3	HQR	102	C2	Unwilling/Complaining
		ASD30	45	2	1.2	CP	*	H	Thirsty
		ASD31	49	7	12	CP	*	H	Unwilling/Complaining
TD		TD25	24	5	6	HQR	NA	C4	Unwilling/Complaining
		TD26	25	2	4.4	HQR	NA	C5	Scared
		TD27	29	5	5	HQR	NA	C5	Scared
		TD28	33	2	2.1	CP	NA	H	Complaining
		TD29	45	16	11	HQR	NA	H	Unwilling/Complaining
		TD30	50	6	7	HQR	NA	H	Complaining
		TD31	51	2	0.7	CP	NA	H	Unwilling

All female subjects were used in the test set.
TD Typically Developed, **ASD** with Autism Spectrum Disorder, **HQR** High Quality Recorder, **CP** Cell Phone, **C1-C2** autism Centers, **C4-C5** health Centers, **H** Home, **NA** Not Applicable
*GARS grades at the time of diagnosis or starting the intervention was not available

Data collection and preprocessing

As mentioned earlier, the data was recorded using high-quality devices and typical smart phones. The high-quality devices were a UX560 Sony voice recorder and a Sony UX512F voice recorder. To use typical smart phones, a voice recording and archiving application was developed and used on various types of smartphones. The application was implemented in such a way that the voices recorded by all devices had a similar format as the voice recorded by the high-quality recorders.

The reason for using various devices was to have a device independent model. Similarly, the place of recording was not restricted to one place to make the results applicable to all places.

The parents and trained voice collectors were asked to record the voices in a quiet environment. Furthermore, there were asked to keep the recorders or smart phones about 25 cm from the subjects' mouth. Despite the proposed two recommendations, there were recorded voices which did not follow the recommendations, and did not have the required quality, so we had to eliminate

them from the study. The reason for crying, was recorded by asking parents, caregivers, or therapist. All voices were recorded in wav format, 16 bits, and with the sampling rate of 44.1 kHz. The cry voices were not due to pain and they were mostly caused by uncomfortable situations children were in. For example, in the health centers, the cries were due to obligations in weight assessment sessions, or in autism centers, they were due to obligations imposed by the first sessions of therapy.

After data collection, there was a preprocessing phase in which only pure cry parts of the recordings, with no other types of vocalization, were selected. From the selected cries, the beginning and the end which contained voice rises and fades were removed to keep only the steady parts of the cries. Also, the uvular/guttural parts of the cries were removed. There were no other preprocess or inspection of the raw data. Each remaining continuous segment of the cries was considered and used as a sample (instance) in this study. Finally, the samples that were too short to be comprehended by our audition were eliminated. In this study, the final samples were between 320 milliseconds to 3 seconds.

The proposed subset instance classifier

To explain the proposed classifier, let's assume that there is a target group of subjects that we want to distinguish from the rest of the subjects, that from now on we call them the rest. Furthermore, each subject in the target group may have several instances that may be used to distinguish the target group from the rest. Fig 1.a shows a situation in which all instances of all subjects of the target group are differentiable using common classifiers that we call Whole Set Instance (WSI) classifiers. In this figure, the circles represent our target group and the triangles represent the rest. The color coding is used to differentiate between the instances of each subject among each group.

In contrast to the situation in Fig. 1.a, in Fig. 1.b the target group cannot be easily distinguished from the rest. In such a situation, there are instances of two subjects in the target group, i.e. red and brown circles that are not easily separable from the instances in the rest (Case 1). Furthermore, there is a subject with no instances, i.e. the orange circles, easily separable from the rest (Case 2). An example of case 1 is tip toe walking in children with ASD, which is common in about 25% of these children who do it most of the times. An example of case 2 are children with ASD who do not tip toe walk. In other words, there are children with ASD who cannot be distinguished from TD children using tip toe walking behavior.

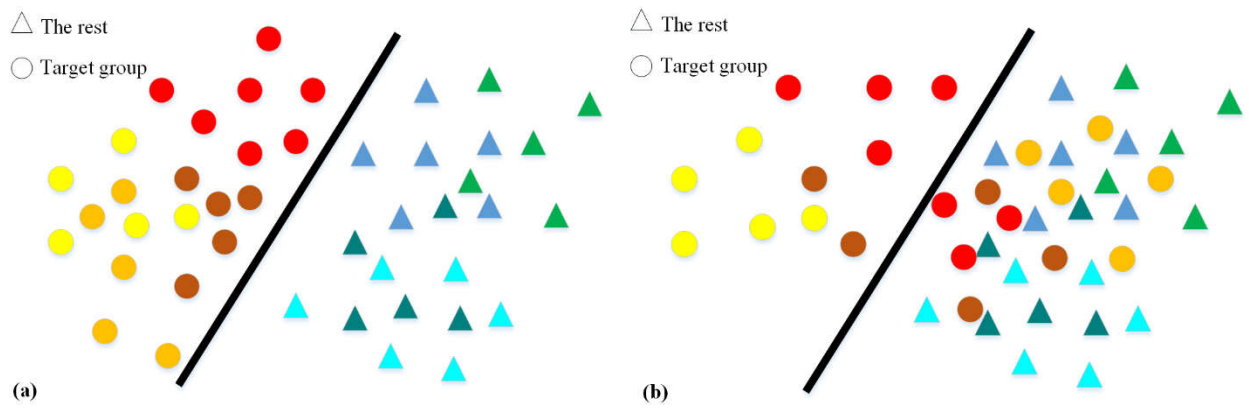

Fig 1. Two different hypothetical types of two-dimensional data of target group and the rest. The instances shown by warm-colored circles and cool-colored triangles are for the target group and the rest, respectively. All instances of a subject have the same color. In (a), all the target group subjects' instances are distinguishable using a classifier. In (b), only some instances of the target group subjects are separable from other instances by a classifier.

Applying any WSI classifier may fail for the data type shown in Fig 1.b. Consequently, we proposed SubSet Instance (SSI) classifier that first finds differentiable instances and then trains a classifier on these instances. As an example, the proposed SSI classifier first tries to find the circles

on the left of the line in Fig. 1.b, using a clustering method. Then, it uses these circles, as *exclusive instances* having a specific feature common in a subset of the target group, to train a classifier separating a subset of the target group.

The steps of common WSI classifiers are shown in Fig 2.a. The steps of our proposed SSI classifier are shown in Fig 2.b. In the SSI classification approach, after the feature extraction and clustering steps, for each cluster a classifier is trained to separate its exclusive instances from the instances of the rest of subjects. In the testing phase, any subject with only one instance classified in the target group (positive instance), is classified in the target group subject. The pseudo code for the proposed approach is given in Algorithms 1 and 2.

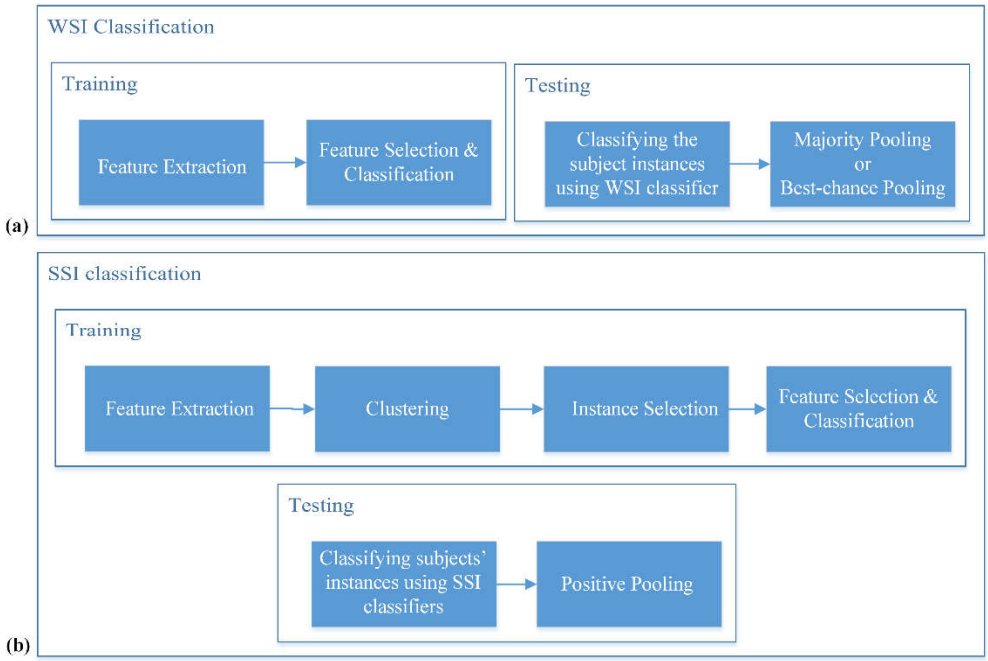

Fig 2. An overall view of WSI and SSI methods. (a) In WSI method, after feature extraction a classifier is trained on all instances and majority pooling (MP) is usually used in testing phase. In this study Best-chance threshold Pooling (BP), that is a threshold-based pooling with the threshold giving the best accuracy on the test set, is also used in order to give the best chance to WSI classifier. (b) In the proposed SSI classifier, after feature extraction, clustering is applied to find and select exclusive instances containing instances of the target group subjects only. Then classifiers are trained using exclusive instances, and a subject is classified in the target group in testing phase if any classifier detects a positive instance for it.

Algorithm 1. Training SSI classifiers

T: set of all target group instances

R: set of all the rest instances

F: set of all Classifiers

ρ: threshold for the number of samples in a cluster

s: the number of minimum samples needed in a cluster to be able to train a classifier for it

C_j: The jth cluster

n: number of clusters

$F = \emptyset$

While $\exists j \, |C_j| > \rho$; while there is a cluster bigger than a threshold or n =1

    $n = n + 1$; increase the number of clusters

    Cluster the $T + R$ into n clusters $C_j, j = 1, \dots, n$

    $EC = \{C_j \subset T\}$; the set of clusters of only exclusive instances, i.e. exclusive clusters

    If $EC \neq \emptyset$; check if there is any exclusive cluster

        For all C_j in EC with $|C_j| > s$

            Train a classifier using positive labels c $\in C_j$ and negative labels $r \in$ R

            Add the classifier to F

        $T = T - \sum_{C_j \subset EC} C_j$; remove the instances of the exclusive clusters from target group instances

        $n = 1$; set 1 to re-start clustering in two groups on the remaining instances

Algorithm 2. Testing SSI classifiers

F: set of trained classifiers

A: set of subject instances

For all instances a of A

    $P = \{a \in A | \exists f, classifies\ a\ as\ positive\ instance\}$

    If $P \neq \emptyset$

        Subject is from target group

    Else

        Subject is from the rest

In the proposed training algorithm of the SSI approach, the goal is to find clusters containing the ASD instances only. Then a classifier is trained using the instances of these clusters and added to

a list of all trained classifiers (lines 7 and 8 of Algorithm 1). As shown in the loop of the algorithm, the data is clustered starting with two clusters. Then the number of clusters is increased till a cluster containing only the target group instances is appeared. The exclusive instances in such a cluster are removed from the set of all target group's instances, and the loop is restarted. Before restarting the loop if the number of instances in this cluster is more than a threshold, a new classifier using these instances is trained and this classifier is added to the set of all trained classifiers. The loop stops when the number of samples in each cluster is less than a threshold.

For testing the subjects using the trained classifiers, all the instances of each subject are classified one by one using all the trained classifiers (line 2 of Algorithm 2). A subject would be classified in the target group if at least one of its instances is classified in the target group at least by one of the classifiers (lines 3 and 4 of Algorithm 2). Otherwise, if there is no instance classified among the target group, the subject is classified as the rest (lines 5 and 6).

Feature extraction

Previous studies working on voice features for discriminating ASD children use different sets of features. These methods share some common features like F0, i.e. the fundamental frequency of a voice, and Mel-Frequency Cepstral Coefficients (MFCC), i.e. coefficients which are representations of the short-term power spectrum of a sound. F0 has been one of the most common features used [15, 16, 22]. However, since age is an important factor affecting F0 [28], this feature is useful when subjects have similar age. MFCC coefficients and several related statistical values have been reported to be useful features in several studies [19, 24, 29]. Considering useful features reported in the previous studies and the specifications of the current study, several features were selected to be used in this work that are explained in the following. .

In this study, we used several features proposed by Motlagh, Moradi and Pouretemad [24] and by Belalcázar-Bolaños et al. [30]. Also, we added a few other promising features. The features used in Motlagh, Moradi and Pouretemad [24] include certain statistics like mean and covariance of voice segments' basic features, like MFCC coefficients. Other statistical features used in it are the mean and variance of frame-wise temporal derivative of the basic features. Frame-wise temporal derivative means subtracting the value of a frame from the value of the next frame. We modified the spectral flatness features by including the range of 125-250 Hz in addition to 250-500 Hz. Each range is divided into 4 octaves and spectral flatness is computed for those octaves. We removed all uninformative and noisy features of the set, i.e. 43% of them, which are explained in the following. The mean of frame-wise temporal derivative of the basic features is removed because it is not a meaningful feature and is equal to taking the difference between the value of the last and the first frames. There are features related to the energy such as the audio power, total loudness, SONE, and the first coefficient of MFCC. The means of these features are omitted to make the classifier independent of the loudness/power in children's voices. Zero crossing rate (ZCR) is omitted too, due to its dependency on the noise in the environment.

The second set of features used in this study is from Belalcázar-Bolaños et al. [30] because it has phonation features, like jitter and shimmer. Jitter and shimmer, which have been reported to be discriminative for ASD, are linked to perceptions of breathiness, hoarseness, and roughness [31]. Other features used from Belalcázar-Bolaños et al. [30] include glottal features related to vocal quality and the closing velocity of the vocal folds [18]. The mean of logarithmic energy feature is omitted by the same reason as other energy-related features.

Finally, to compute the features, each instance was divided into 20 milliseconds frames.

Results

In this part, the performance of our proposed SSI classifier against common WSI classifier is evaluated on a test set of ASD and TD subjects. Each subject has multiple instances which are cleaned using the criteria explained in the Data collection and preprocessing section. The subjects which had at least one accepted instance are used in the training and testing phases, which are shown in Tables 1 and 2.

The details of the training phase for the common WSI approach and the proposed SSI approach are explained in the Appendix for the details). The output of the SSI approach is two classifiers, each of them works by thresholding on a feature. These features are Variance of Frame-wise Temporal Derivative (VFTD) of the 7th MFCC coefficient and VFTD of the 5th SONE coefficient (Fig 3).

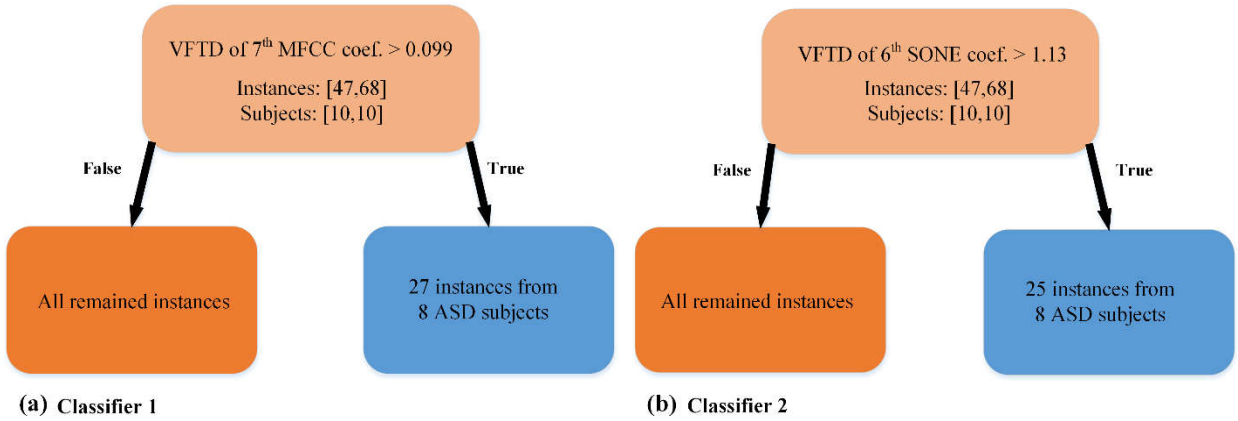

Figure 3 Two classifiers trained on two exclusive clusters found during SSI classifier training phase. (a) Variance of Frame-wise Temporal Derivative (VFTD) of the 7th MFCC coefficient separates 27 instances of 8 ASD subjects from all TD instances of the training set. (b) VFTD of the 5th SONE coefficient separates 17 instances of 7 ASD subjects from all TD instances of the training set.

The number of instances of ASD subjects in the training set correctly detected by the first and second classifiers is shown in the first and second columns of Table 3, respectively. The best-

resulting classifier for the WSI approach is Radial Basis Function-Support Vector Machine (RBF-SVM).

Table 3.The number of instances of each subject in the training set that are classified as ASD using each trained SSI classifier.

ID	First SSI classifier	Second SSI classifier
ASD1	8	3
ASD2	1	2
ASD3	3	1
ASD4	10	9
ASD5	0	0
ASD6	1	3
ASD7	1	0
ASD8	1	2
ASD9	0	1
ASD10	2	4

The following subsection includes the comparative results of the two classification approaches.

Classification results

The classification results on the test set for different classifiers are shown in Table 4. The portion of each subject's instances correctly classified by each classifier is written in percentage under the name of the classifier. The decision made by WSI and SSI classifiers for each subject is shown by A and T for being ASD or TD. To classify each subject using WSI classifier Majority Pooling (MP) and Best-chance threshold Pooling (BP), which is a threshold-based pooling with the threshold giving the best accuracy on the test set for male subjects, approaches are used. For the male subjects, MP has specificity, sensitivity, and precision equal to 100%, 35.71%, and 67.85%, respectively. BP leads to specificity, sensitivity, and precision equal to 85.71%, 71.42%, and 78.57%, respectively. The threshold for BP was set to 20% that means if 20% of instances of a subject was classified as ASD instance, the subject was classified as ASD. The results of the

percentage of instances correctly classified by the two classifiers in the SSI approach, are shown as C_1 (the first SSI classifier) and C_2 (the second SSI classifier) in Table 4. The aggregated result is the final decision of the SSI classifier which is shown in the decision column under the SSI classification section. Achieved specificity, sensitivity, and precision using the proposed method for the male subjects are 100% , 85.71%, and 92.85%, respectively.

Table 4.The results of classifiers (WSI and SSI) on the instances of each subject in the test set.

	Typically developed children							Children with autism						
	ID	Portion of instances classified as TD in percentage and the decision						ID	Portion of instances classified as ASD in percentage and the decision					
		WSI classification			SSI classification				WSI classification			SSI classification		
		SVM	Dec.		C_1	C_2	Dec.		SVM	Dec.		C_1	C_2	Dec.
			MP	BP						MP	BP			
Male subjects	TD11	100	T	T	100	100	T	ASD11	50	A	A	17	50	A
	TD12	100	T	T	100	100	T	ASD12	33	T	A	11	28	A
	TD13	100	T	T	100	100	T	ASD13	33	T	A	33	0	A
	TD14	100	T	T	100	100	T	ASD14	20	T	A	20	20	A
	TD15	100	T	T	100	100	T	ASD15	0	T	T	0	40	A
	TD16	100	T	T	100	100	T	ASD16	50	A	A	100	0	A
	TD17	100	T	T	100	100	T	ASD17	0	T	T	0	100	A
	TD18	83	T	T	100	100	T	ASD18	50	A	A	50	50	A
	TD19	100	T	T	100	100	T	ASD19	0	T	T	0	0	T
	TD20	80	T	A	100	100	T	ASD20	42	T	A	42	16	A
	TD21	100	T	T	100	100	T	ASD21	100	A	A	0	0	T
	TD22	100	T	T	100	100	T	ASD22	0	T	T	0	50	A
	TD23	75	T	A	100	100	T	ASD23	33	T	A	33	17	A
	TD24	92	T	T	100	100	T	ASD24	86	A	A	86	86	A
	Acc. %		100	85.71			100			35.71	71.42			85.71
Female subjects	TD25	100	T	T	100	100	T	ASD25	42	T	A	17	0	A
	TD26	100	T	T	100	100	T	ASD26	60	A	A	60	20	A
	TD27	100	T	T	100	100	T	ASD27	50	A	A	0	0	T
	TD28	100	T	T	100	100	T	ASD28	100	A	A	0	50	A
	TD29	100	T	T	100	100	T	ASD29	62	A	A	50	50	A
	TD30	67	T	A	100	100	T	ASD30	100	A	A	50	50	A
	TD31	100	T	T	100	100	T	ASD31	0	T	T	0	0	T
	Acc. %		100	85.71			100			71.42	85.71			71.42

Each classifier result on a subject's instances are reported in percentage.
Dec. Decision, **MP** Majority Pooling, **BC** Best-chance threshold Pooling, **C1** Classifier1, **C2** Classifier2, **A** Autistic, **T** Typical Developing, **Acc.** Accuracy

To further show the applicability of the proposed approach on the female subjects, we applied the male-subjects-only-trained classifiers on the test set of the female subjects. The results are shown in the last row of Table 4 which show that the MP approach has specificity, sensitivity, and precision equal to 100%, 71.42%, and 85.71%, respectively. Furthermore, the BP approach gives specificity, sensitivity, and precision all equal to 85.71%, respectively, for female subjects. The results of proposed SSI classifier were 100% specificity, 71.42% sensitivity, and 85.71% precision.

A two-dimensional scatter plot of the two features, used in C_1 and C_2 classifiers, are shown in Fig 4. As can be seen in this figure, the instances of an ASD subject are scattered in the area containing instances of both TD and ASD subjects. Nevertheless, there are some instances for this subject uniquely distinguishable using the selected two features.

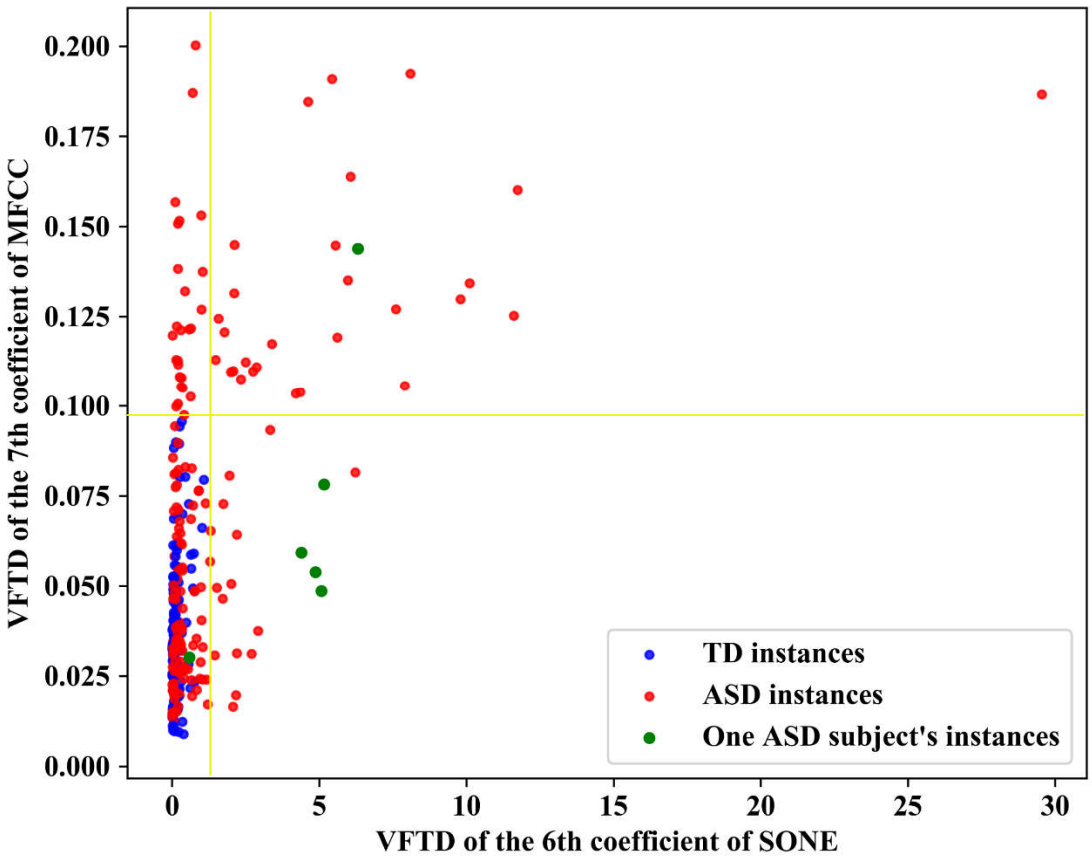

Fig 4. Instances of several ASD and TD subjects scattered in the space of two features given by the proposed SSI method. VFTD stands for Variance of Frame-wise Temporal Derivative. The instances of a chosen ASD subject are illustrated in green to show that a subject may have instances in the area common with TD instances besides those two areas separated by thresholds as ASD. The mentioned ASD subject (with green instances) is tagged as ASD due to having at least one instance with the greater value than at least one of the thresholds on the two features.

We compared our proposed method with the only method [24] on the cry features (Table 5). The results show the superiority of our method compared to the previously proposed method.

Table 5.Comparison of the results on the test set using the two methods; SSI approach and a baseline approach.

		Sensitivity	Specificity	Precision
Male subjects	SSI	85.71%	100%	92.85%
	Baseline	50.58%	81%	65%
Female subjects	SSI	71.42%	100%	85.71%
	Baseline	21%	86.48%	53%

Discussion and conclusion

In this paper, we presented a novel cry-based screening method to distinguish between children with autism and typically developed children. In the proposed method, the groups of children with autism who have specific features in their cry voices can be determined. This method is based on a new classification approach called SubSet Instance (SSI) classifier. A nice property of the proposed SSI classifier, in the case of voice-based autism screening, is its high specificity that a normal child can be detected with no error.

We applied the proposed method on a group of subjects consisted of 31 children with ASD between 20 and 53 months of age and 31 TD children between 18 and 51 months of age. The two features, found in this study, was used to train a classifier on 10 male subjects with ASD and 10 male TD subjects. Then, the classifier was used to distinguish 14 male subjects with ASD from 14 male TD subjects, reaching 92.8% accuracy. Although, the trained classifier based on male children had sufficient accuracy, the classification rate for female subjects was about 7% lower. In the future, we have to increase the number of female subjects to determine whether a single classifier can be used for both male and female subject.

It should be mentioned that our train and test data were completely separate making the trained model more general. The features found in this study, were applicable in a range of ages. This is in contrast to other approaches that either used a dataset of children with a specific age [18, 19] or used age information for classification [17]. Due to the age invariant features found in this study, it can be claimed that there are markers in the voice of children with ASD that are sustained in a range of ages at least.

To the best of our knowledge, [17] and [19] were the only studies on screening children with autism using voice features on children younger than 2 years old. Our proposed method has better precisions than these two, i.e. 6% more than [17] and 17% more than [19], using only cry features. Using cry features as suitable biomarkers for autism screening matches the claimed in [21].

Finally, the proposed approach was also tested on 57 subjects, 56 TD children and one who diagnosed as Unspecified Neuro Developmental Disorder (UNDD), between10 to 18 months of age. The classifier distinguished the UNDD child from the rest. This suggests that a) the system can be used for children under 2 years of age, and b) it may be able to distinguish more general UNDD children from TD children.

In the future, we would need to get more data to validate these results with more confidence. We believe that this approach can be used to perform early screening under 18 months of age. Thus, we need to collect data and test the approach on the data of children under 18 months of age. Finally, we have to check the proposed approach and extracted features on other neurodevelopmental disorders, such as attention deficit hyperactivity disorder, to evaluate the capability of the approach to distinguish the children with these disorders from TD children.

Acknowledgements

We would like to thank the Center for Treatment of Autism Disorder (CTAD) and its members for supporting this study. Also, we like to thank every family who helped to collect the cry voices of their children.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Ethical approval for this project was obtained by the ethical committee of Shahid Beheshti University of Medical Sciences and Health Services.

References:

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub; 2013.
2. Chen JL, Sung C, Pi S. Vocational rehabilitation service patterns and outcomes for individuals with autism of different ages. J Autism Dev Disord. 2015;45(9):3015-29.
3. Fakhoury M. Autistic spectrum disorders: A review of clinical features, theories and diagnosis. Int J Dev Neurosci. 2015;43:70-7.
4. Constantino JN, Charman T. Diagnosis of autism spectrum disorder: reconciling the syndrome, its diverse origins, and variation in expression. Lancet Neurol. 2016;15(3):279-91.
5. Calderoni S, Billeci L, Narzisi A, Brambilla P, Retico A, Muratori F. Rehabilitative interventions and brain plasticity in autism spectrum disorders: focus on MRI-based studies. Front Neurosci. 2016;10:139.
6. Brentani H, Paula CSd, Bordini D, Rolim D, Sato F, Portolese J, et al. Autism spectrum disorders: an overview on diagnosis and treatment. Braz J Psychiatry. 2013;35:S62-S72.
7. Mandell DS, Novak MM, Zubritsky CD. Factors associated with age of diagnosis among children with autism spectrum disorders. Pediatrics. 2005;116(6):1480-6.
8. Zachor DA, Itzchak EB. Treatment approach, autism severity and intervention outcomes in young children. Res Autism Spectr Disord. 2010;4(3):425-32.
9. Thabtah F, Peebles D. Early Autism Screening: A Comprehensive Review. Int J Environ Res Public Health. 2019;16(18):3502.
10. Emerson RW, Adams C, Nishino T, Hazlett HC, Wolff JJ, Zwaigenbaum L, et al. Functional neuroimaging of high-risk 6-month-old infants predicts a diagnosis of autism at 24 months of age. Sci Transl Med. 2017;9(393):eaag2882.
11. Bosl WJ, Tager-Flusberg H, Nelson CA. EEG analytics for early detection of autism spectrum disorder: a data-driven approach. Sci Rep. 2018;8(1):6828.12. Momeni N, Bergquist J, Brudin L, Behnia F, Sivberg B, Joghataei M, et al. A novel blood-based biomarker for detection of autism spectrum disorders. Translational psychiatry. 2012;2(3):e91.
12. Momeni N, Bergquist J, Brudin L, Behnia F, Sivberg B, Joghataei M, et al. A novel blood-based biomarker for detection of autism spectrum disorders. Transl Psychiatry. 2012;2(3):e91.
13. Glatt SJ, Tsuang MT, Winn M, Chandler SD, Collins M, Lopez L, et al. Blood-based gene expression signatures of infants and toddlers with autism. J Am Acad Child Adolesc Psychiatry. 2012;51(9):934-44.
14. Croen LA, Braunschweig D, Haapanen L, Yoshida CK, Fireman B, Grether JK, et al. Maternal mid-pregnancy autoantibodies to fetal brain protein: the early markers for autism study. Biol Psychiatry. 2008;64(7):583-8.
15. Brisson J, Martel K, Serres J, Sirois S, Adrien JL. Acoustic analysis of oral productions of infants later diagnosed with autism and their mother. Infant Ment Health J. 2014;35(3):285-95.
16. Nakai Y, Takiguchi T, Matsui G, Yamaoka N, Takada S. Detecting abnormal word utterances in children with autism spectrum disorders: machine-learning-based voice analysis versus speech therapists. Perceptual and motor skills. 2017;124(5):961-73.
17. Oller D, Niyogi P, Gray S, Richards J, Gilkerson J, Xu D, et al. Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. Proc Natl Acad Sci. 2010;107(30):13354-9.

18. Santos JF, Brosh N, Falk TH, Zwaigenbaum L, Bryson SE, Roberts W, et al. Very early detection of autism spectrum disorders based on acoustic analysis of pre-verbal vocalizations of 18-month old toddlers. Proc IEEE Int Conf Acoust Speech Signal Process; 2013; Vancouver, BC, Canada. IEEE; 2013. p. 7567-7571
19. Pokorny FB, Schuller BW, Marschik PB, Brueckner R, Nyström P, Cummins N, et al. Earlier Identification of Children with Autism Spectrum Disorder: An Automatic Vocalisation-Based Approach. INTERSPEECH; 2017. p. 309-313
20. Eyben F, Scherer KR, Schuller BW, Sundberg J, André E, Busso C, et al. The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. IEEE Transactions on Affective Computing. 2015;7(2):190-202.
21. Esposito G, Hiroi N, Scattoni ML. Cry, Baby, Cry: Expression of Distress As a Biomarker and Modulator in Autism Spectrum Disorder. Int J Neuropsychopharmacol. 2017;20(6):498-503.
22. Sheinkopf SJ, Iverson JM, Rinaldi ML, Lester BM. Atypical Cry Acoustics in 6-Month-Old Infants at Risk for Autism Spectrum Disorder. Autism Res. 2012;5(5):331-9.
23. Orlandi S, Manfredi C, Bocchi L, Scattoni ML, editors. Automatic newborn cry analysis: a non-invasive tool to help autism early diagnosis. 2012 Conf Proc IEEE Eng Med Biol Soc; 2012; San Diego, CA, USA. IEEE. 2012. p. 2953-6
24. Motlagh SHRE, Moradi H, Pouretemad H, editors. Using general sound descriptors for early autism detection. Control Conference (ASCC), 2013 9th Asian; 2013; Istanbul, Turkey. IEEE. 2013. p. 1-5
25. Barrow WJ, Jaworski M, Accardo PJ. Persistent toe walking in autism. J Child Neurol. 2011;26(5):619-21.
26. Gilliam JE. Gilliam autism rating scale: GARS 2: Pro-ed; 2006.
27. Sadeghi S, Pouretemad H, Khosrowabadi R, Fathabadi J, Nikbakht S. Behavioral and electrophysiological evidence for parent training in young children with autism symptoms and excessive screen-time. Asian J Psychiatry. 2019;45:7-12. doi: https://doi.org/10.1016/j.ajp.2019.08.003.
28. Esposito G, Venuti P. Developmental changes in the fundamental frequency (f0) of infants' cries: a study of children with Autism Spectrum Disorder. Early Child Dev Care. 2010;180(8):1093-102.
29. Marchi E, Schuller B, Baron-Cohen S, Golan O, Bölte S, Arora P, et al. Typicality and emotion in the voice of children with autism spectrum condition: Evidence across three languages. INTERSPEECH; 2015. p. 115-119
30. Belalcázar-Bolaños EA, Orozco-Arroyave JR, Vargas-Bonilla JF, Haderlein T, Nöth E, editors. Glottal Flow Patterns Analyses for Parkinson's Disease Detection: Acoustic and Nonlinear Approaches. Proceedings of the International Conference on Text, Speech, and Dialogue; 2016; Cham: Springer; 2016.
31. Bone D, Lee C-C, Black MP, Williams ME, Lee S, Levitt P, et al. The psychologist as an interlocutor in autism spectrum disorder assessment: Insights from a study of spontaneous prosody. J Speech Lang Hear Res. 2014;57(4):1162-77.

Appendix

In this part, we give a short overview of the classification approaches used in the comparison. Afterwards, we explore the details of trained SSI classifier.

Classification approaches

We are using two types of classifiers in this paper as explained in the Proposed method Section. The first type of classifiers is commonly used WSI classifier. The other classifier is our proposed SSI classifier. We give the details of classifiers in the following.

WSI Classifiers:

We tested several common WSI classifiers, but we report only the result of SVM with RBF kernel and with no feature selection, which gives the best average accuracy. It should be noted that several feature selection approaches, like L1-SVM and backward elimination, were tested but they only worsened the results. We use group 5-fold cross-validation for tuning hyper-parameters. Group K-fold means that all instances of each subject are placed in only one of the folds. This prevents having the same subject's instances in the train and validation folds simultaneously. In each fold, there are two ASD and two TD subjects.

Two approaches were exploited to combine the decisions on different samples of a subject in WSI approach. The first approach is majority pooling which classifies a subject as ASD if the number of instances classified as ASD was more than 50 percent of all instances. The second approach is threshold-based pooling which is similar to the first approach except a threshold other than 50 percent is used.

SSI Classifiers:

Our proposed SSI algorithm was explained in Algorithm 1. The agglomerative clustering is used in this algorithm. Before applying the algorithm, we balance the number of instances of two groups by upsampling. Decision tree is used for the classification used in Algorithm 1. Decision tree is chosen with the hope that only a few features can be chosen to make the classification possible. The threshold r, which is the number of minimum samples needed in a cluster to be able to train a classifier for it, in the algorithm is set to 10.

Training SSI classifier:

After running Algorithm 1 on our data, two exclusive clusters with enough (at least 10 in this study) instances were found, and therefore two classifiers were trained corresponding to each of them. One of this exclusive clusters had 11 instances from 4 ASD subjects in Table 1. These 11 instances were composed from 6 out of 9 instances of subject ASD1, 2 out of 4 instances of subject ASD10, 1 out of 2 instances of subject ASD8 and 2 out of 4 instances of subject ASD6. As explained in the algorithm, the classifier, which was a decision tree for each cluster, was trained using the ASD instances versus TD instances. Interestingly, only one feature was enough to discriminate instances in the cluster from all TD instances. Among those features that can discriminate the cluster's instances, we select a feature which can discriminate more ASD subjects from the set of all subjects with a simple threshold. The classifier obtained by thresholding on this feature is the first classifier. The feature is Variance of Frame-wise Temporal Derivative (VFTD) of the 7th MFCC coefficient. This feature supports our expert's report saying variations in the cry of ASD children is more than TD children. From 10 ASD children, 8 of them can be discriminated using this feature. For each subject, the number of instances found by this classifier is shown in the 2nd column of Table 2.

After excluding the samples from the first classification, the second exclusive cluster with the similar method explained in the previous paragraph gives the second classifier. This cluster includes all instances of subject ASD4. Again for this cluster, a threshold on only one feature can discriminate all instances of ASD children and the feature for this classifier is VFTD of the 6th SONE coefficient. SONE is a unit of loudness which is a subjective perception of sound pressure. Having higher VFTD of this feature in ASD group confirms the experience by the experts mentioned before. Among all ASD subjects, eight have instances with VFTD of the 6th SONE higher than a threshold (Shown in the 3rd column of Table 3). The results of classification based on these two features are depicted in Fig 3. As it is mentioned in the proposed method section, the subjects with at least one instance classified, by a classifier, as ASD would be considered as an ASD subject during the test time.